1. **Basic Information**

Project title: Indian Premier League Archive Explorer
Names: Harsh Mahajan, Kunal Manjare, Dhruvil Shah
Email IDs: u1413898@utah.edu, u1419704@utah.edu, u1420007@utah.edu
uID: u1413898, u1419704, u1420007
Link to repository: https://github.com/kunal911/Data-Visualization-Final-Project-

2. **Background and Motivation**

Cricket is the second most played sport with over 2.6 billion viewers across the world and the Indian Premier League (IPL) is considered to be the pinnacle of franchise cricket contributing to over 220 million viewers. Before we dive into our motivation for this project, we would like to talk a bit about cricket and IPL.

Cricket

There are primarily three types of international cricket matches-- test matches, one day internationals, and Twenty20. Among these, one day internationals (ODIs) are the most commonly played type of matches in international tournaments including the ICC Cricket World Cup and the ICC Champions Trophy. Matches are played between two teams each consisting of 11 players.

1. Each One Day International(ODI) match consists of two innings requiring both teams to bat and field once. Each inning is composed of 50 overs and each over consists of six deliveries bowled by the one player from the fielding team. The goal of the batting team is to score as many runs as possible and avoid their player's being dismissed by the fielding team while doing so. At the end of the match, the team with a higher score wins.
2. Twenty20 cricket, also called T20, truncated the form of cricket that revolutionized the game when it was introduced in 2003.The basic rules are the same as for the longer versions, but innings are limited to 20 overs a side, with a maximum of four overs for each bowler and restrictions on the placement of fielders designed to encourage big hitting by the batsmen and high scores.
3. A Test match is the longest form of cricket and the format in which the real essence of the game lies. It is a true test of the ability and temperament of a cricketer and the most intriguing of all formats. A test match would go on for a duration of five days with two innings per side. A standard day in test cricket consists of three sessions with 30 overs in each session. The breaks between sessions are 40 minutes for lunch and 20 minutes for tea.

IPL

The 2008-founded Indian Premier League (IPL) is a professional Twenty20 (T20) cricket league in India. Major Indian cities have teams in the league, which uses a round-robin group and knockout system. The IPL, which was created by the Board of Control for Cricket in India (BCCI), has grown to be the most lucrative and well-liked cricket league in the world. Matches typically start in the late afternoon or early evening so that at least

some of the games can be played at night under floodlights for international television broadcasts.

With each game more data (like balls faced, runs, wickets etc.) is being generated and with each chunk of data being produced, it becomes harder to keep a track of everything. The struggle of going through multiple files everytime motivated us to work on a solution.

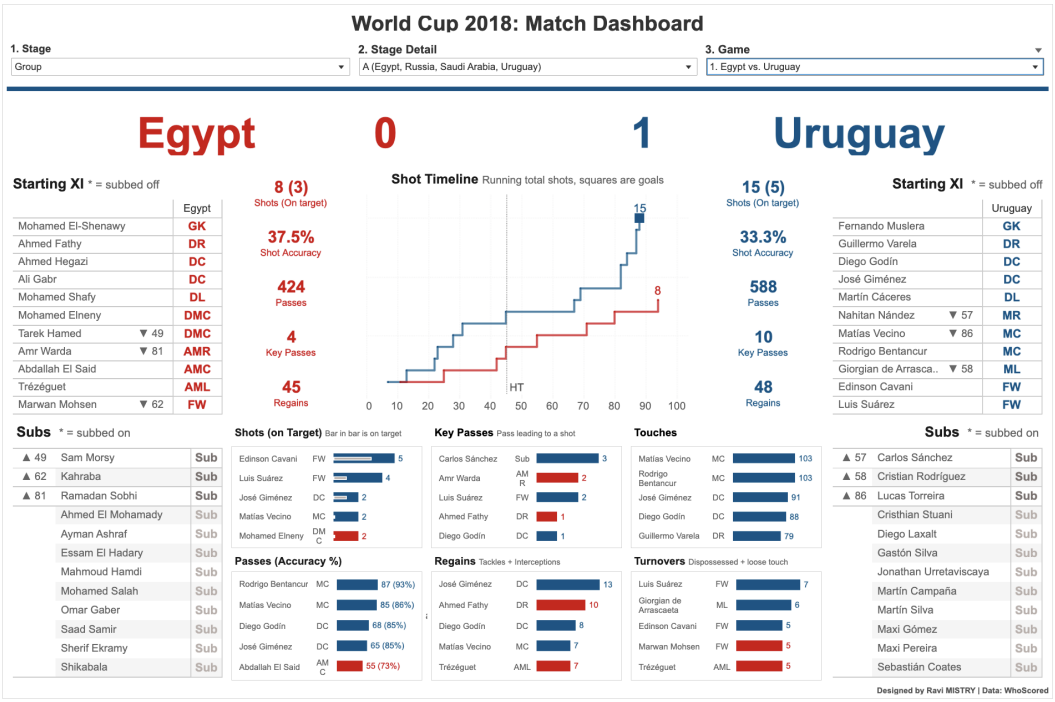
3. **Project Objectives**
   The objective of this project is to combine our skills at creating beautiful visualizations and our knowledge & passion for cricket and turn it into something meaningful. This project is meant for people who would like to dive deep into the archives of IPL and use the data to create their own hypothesis. We are attempting to do this by creating a comprehensive dashboard while following the design guidelines taught to us.

4. **Data**
   The required dataset is present on kaggle and cricsheet websites in CSV and JSON format. For every match the data contains the following information: teams,players involved. It also contains ball-by-ball details such as the bowler, the player who played the delivery, runs scored on that delivery etc.
   https://www.kaggle.com/datasets/vora1011/ipl-2008-to-2021-all-match-dataset?select=IPL_Matches_2008_2022.csv
   https://cricsheet.org/downloads/

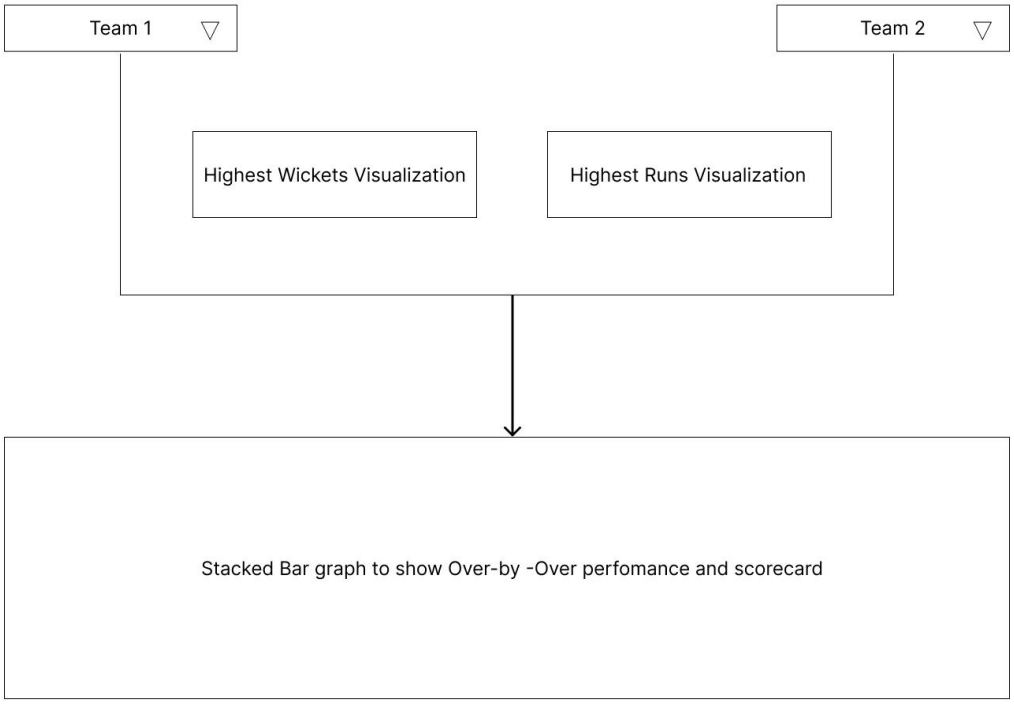
5. **Data Processing**
   The data requires some preprocessing as it contains some information/features which are not required for our project. Also, we will have to perform some data transformation because for implementing some features we will have to parse multiple JSON files to extract and combine information and create our own objects.

   We will be using the combined information from the JSON for the ball by ball visualization while the information from the CSv will be used to for the ranking history of teams.

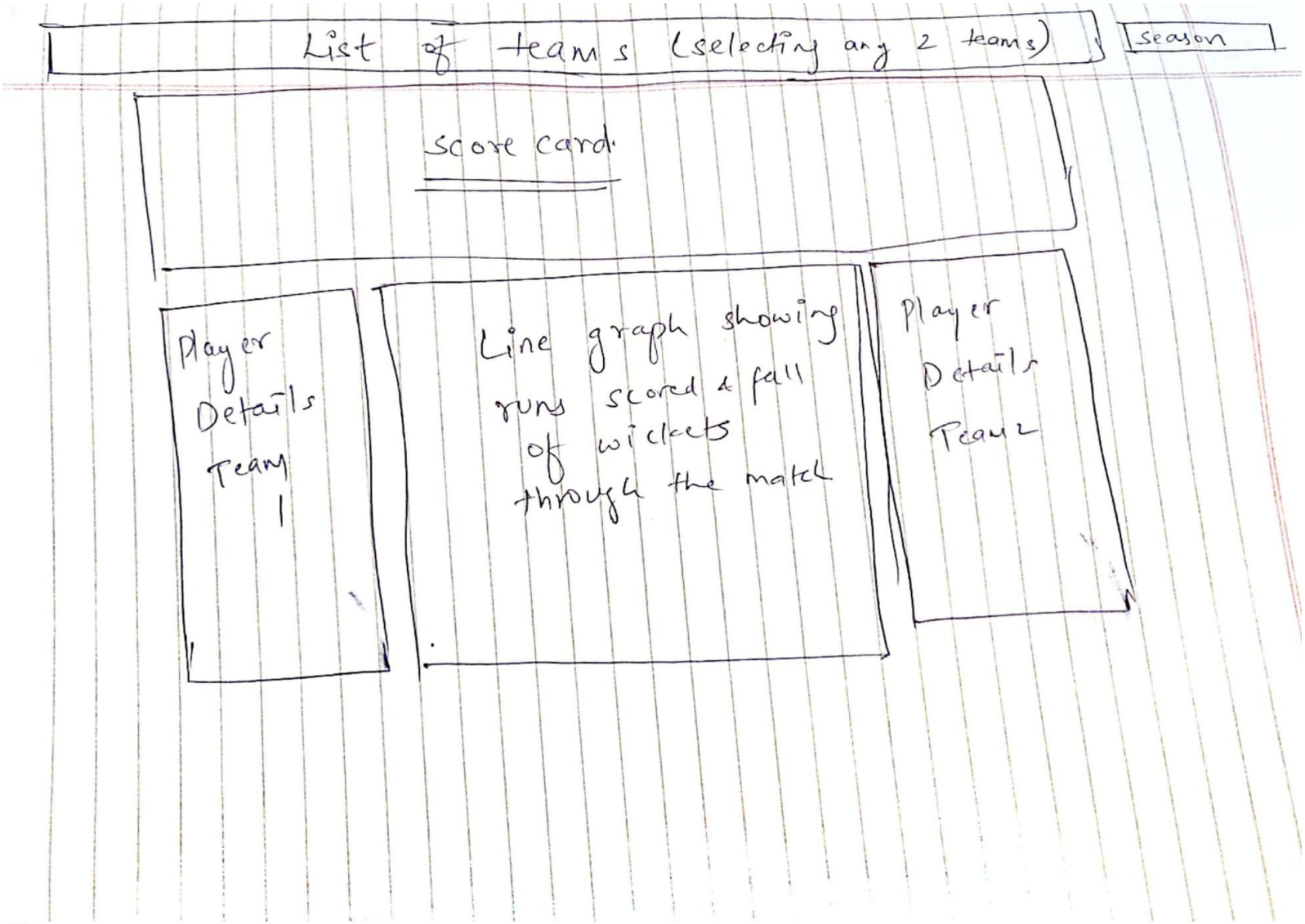
6. **Visualization Design**
   Reference Visualization:


World Cup 2018: Match Dashboard
1. Stage
Group
2. Stage Detail
A (Egypt, Russia, Saudi Arabia, Uruguay)
3. Game
1. Egypt vs. Uruguay
Egypt
0
1
Uruguay
Starting XI * = subbed off
Egypt
Mohamed El-Shenawy GK
Ahmed Fathy DR
Ahmed Hegazi DC
Ali Gabr DC
Mohamed Shafy DL
Mohamed Elneny DMC
Tarek Hamed ▼ 49 DMC
Amr Warda ▼ 81 AMR
Abdallah El Said AMC
Trézéguet AML
Marwan Mohsen ▼ 62 FW
8 (3)
Shots (On target)
37.5%
Shot Accuracy
424
Passes
4
Key Passes
45
Regains
Shot Timeline Running total shots, squares are goals
15
8
HT
0 10 20 30 40 50 60 70 80 90 100
15 (5)
Shots (On target)
33.3%
Shot Accuracy
588
Passes
10
Key Passes
48
Regains
Starting XI * = subbed off
Uruguay
Fernando Muslera GK
Guillermo Varela DR
Diego Godín DC
José Giménez DC
Martín Cáceres DL
Nahitan Nández ▼ 57 MR
Matías Vecino ▼ 86 MC
Rodrigo Bentancur MC
Giorgian de Arrasca.. ▼ 58 ML
Edinson Cavani FW
Luis Suárez FW
Subs * = subbed on
▲ 49 Sam Morsy Sub
▲ 62 Kahraba Sub
▲ 81 Ramadan Sobhi Sub
Ahmed El Mohamady Sub
Ayman Ashraf Sub
Essam El Hadary Sub
Mahmoud Hamdi Sub
Mohamed Salah Sub
Omar Gaber Sub
Saad Samir Sub
Sherif Ekramy Sub
Shikabala Sub
Shots (on Target) Bar in bar is on target
Edinson Cavani FW 5
Luis Suárez FW 4
José Giménez DC 2
Matías Vecino MC 2
Mohamed Elneny DM C 2
Passes (Accuracy %)
Rodrigo Bentancur MC 87 (93%)
Matías Vecino MC 85 (86%)
Diego Godín DC 68 (85%)
José Giménez DC 65 (85%)
Abdallah El Said AM C 55 (73%)
Key Passes Pass leading to a shot
Carlos Sánchez Sub 3
Amr Warda AM R 2
Luis Suárez FW 2
Ahmed Fathy DR 1
Diego Godín DC 1
Regains Tackles + Interceptions
José Giménez DC 13
Ahmed Fathy DR 10
Diego Godín DC 8
Matías Vecino MC 7
Trézéguet AML 7
Touches
Matías Vecino MC 103
Rodrigo Bentancur MC 103
José Giménez DC 91
Diego Godín DC 88
Guillermo Varela DR 79
Turnovers Dispossessed + loose touch
Luis Suárez FW 7
Giorgian de Arrascaeta ML 6
Marwan Mohsen FW 5
Edinson Cavani FW 5
Trézéguet AML 5
Subs * = subbed on
▲ 57 Carlos Sánchez Sub
▲ 58 Cristian Rodríguez Sub
▲ 86 Lucas Torreira Sub
Cristhian Stuani Sub
Diego Laxalt Sub
Gastón Silva Sub
Jonathan Urretaviscaya Sub
Martín Campaña Sub
Martín Silva Sub
Maxi Gómez Sub
Maxi Pereira Sub
Sebastián Coates Sub
Designed by Ravi MISTRY | Data: WhoScored

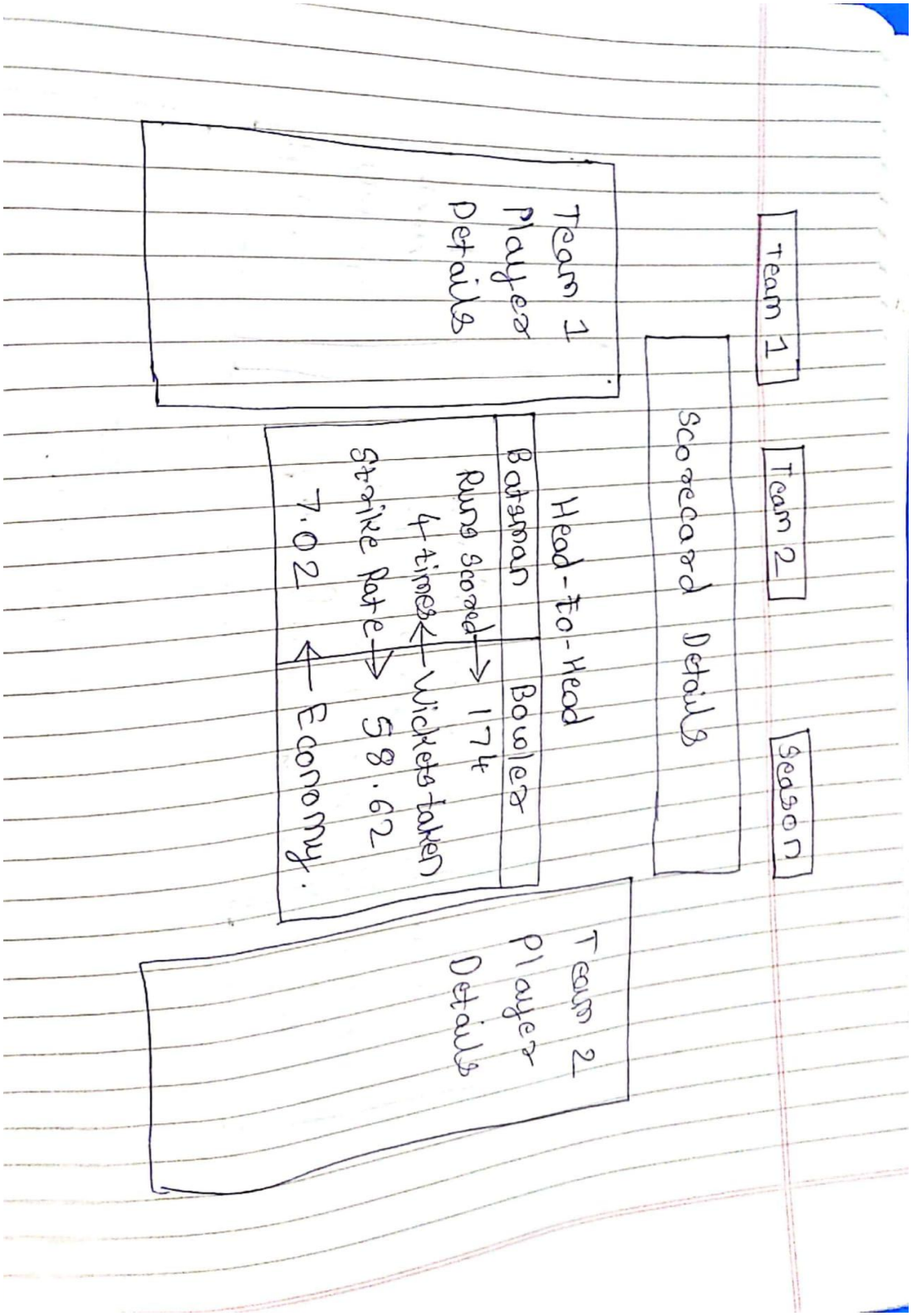
Design 1:


Team 1
Team 2
Highest Wickets Visualization
Highest Runs Visualization
Stacked Bar graph to show Over-by -Over perfomance and scorecard

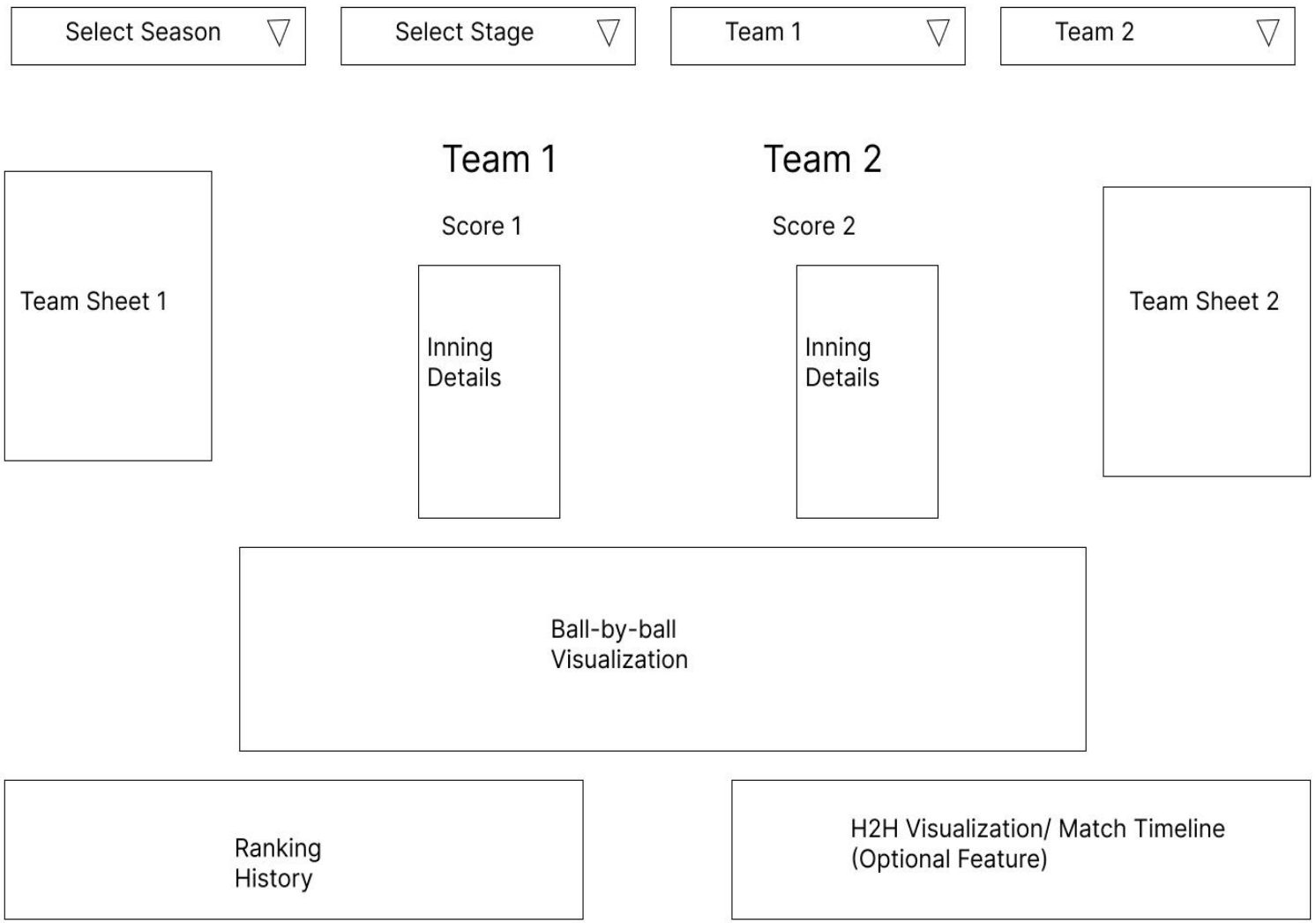
Design 2:



Design 3:

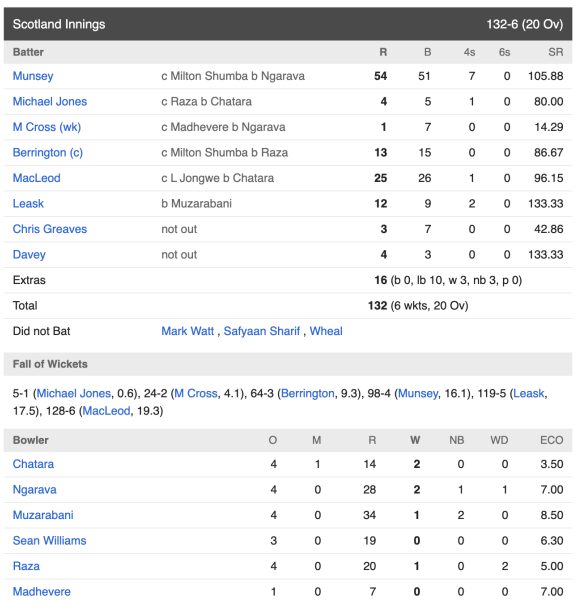

Team 1
Team 2
Season
Scorecard Details
Team 1 Player Details
Head-to-Head
Batsman
Bowler
Runs Scored
174
4 times
Wickets taken
Strike Rate
58.62
7.02
Economy.
Team 2 Player Details

Final Design:



7. Must-have Features

1. **A scoreboard to display in-depth performance of each team:** We plan to use a tabular representation to display the runs scored by each player on the batting team and the figures of the bowling team. This is a representation which is currently being used by a very famous application, we are looking at a way to make the data easily readable at the first glance by using various attributes like colors and bar graphs for batting data and heat maps for bowling data.



Scotland Innings						132-6 (20 Ov)
Batter		**R**	B	4s	6s	SR
Munsey	c Milton Shumba b Ngarava	**54**	51	7	0	105.88
Michael Jones	c Raza b Chatara	**4**	5	1	0	80.00
M Cross (wk)	c Madhevere b Ngarava	**1**	7	0	0	14.29
Berrington (c)	c Milton Shumba b Raza	**13**	15	0	0	86.67
MacLeod	c L Jongwe b Chatara	**25**	26	1	0	96.15
Leask	b Muzarabani	**12**	9	2	0	133.33
Chris Greaves	not out	**3**	7	0	0	42.86
Davey	not out	**4**	3	0	0	133.33
Extras		**16** (b 0, lb 10, w 3, nb 3, p 0)				
Total		**132** (6 wkts, 20 Ov)				
Did not Bat	Mark Watt , Safyaan Sharif , Wheal					

Fall of Wickets

5-1 (Michael Jones, 0.6), 24-2 (M Cross, 4.1), 64-3 (Berrington, 9.3), 98-4 (Munsey, 16.1), 119-5 (Leask, 17.5), 128-6 (MacLeod, 19.3)

Bowler	O	M	R	W	NB	WD	ECO
Chatara	4	1	14	**2**	0	0	3.50
Ngarava	4	0	28	**2**	1	1	7.00
Muzarabani	4	0	34	**1**	2	0	8.50
Sean Williams	3	0	19	**0**	0	0	6.30
Raza	4	0	20	**1**	0	2	5.00
Madhevere	1	0	7	**0**	0	0	7.00

2. **Playing XI(squad) of both teams**

3. **Ball-by-ball representation of the game:** We are looking at multiple various types of visualizations to represent the runs scored on each ball along with the run-rate, a few options are to use a bar chart or a dot chart to show the runs and a line chart or an area chart to depict the run rate. Current available resources use normal text to depict this data which makes it very difficult for the user to search for a particular ball/over.

4. **Ranking history of both teams in the tournament:** IPL has been established for 14 years now and it would be extremely tedious to search for the ranking of a particular team in a particular season. To solve this issue we plan to create a line graph of the way the ranking of a team has changed over a particular season.

Adding these features would help us make a one-stop dashboard for all IPL archives, which can make the life of the user very easy and reduce their efforts and time exponentially.

8. **Optional Features**

We are considering some additional features to include in our final output.

1.Representing stats of each individual player: by most runs scored and also most wickets taken by an individual player.

2.Displaying scoreboard for each season: analysis of highest as well as lowest totals scored by any team in IPL.

3.Past performances: Displaying how many games the teams have played head-to-head matches and the outcome of those games.

4. Additional analysis of unique shots: A visual representation of a player shot and score would additionally help to highlight the player information and would be nice to include.

9. **Project Schedule**

Date	**Task**	**Is it Completed?**
10/03/2022	Form Team	Yes
10/05/2022	Meet to discuss possible topics	Yes
10/07/2022	Submit form with team detail	Yes
10/13/2022	Standup session to discuss various visualization design	Yes
10/14/2022	Critique of rough designs	Yes
10/14/2022	Selection of design and dataset	Yes
10/15/2022	Begin with data preprocessing	Ongoing
10/16/2022	Begin project report	Yes
10/21/2022	Submit project report	Yes
10/28/2022	Divide each visualization among the members	
11/11/2022	Project Milestone	
11/18/2022	Smoothen out kinks and look at implementing must-have features	